Memory-Based In Situ Learning for Unmanned Vehicles

Patrick McDowell, Brian S. Bourgeois, and Donald A. Sofge, Naval Research Laboratory
S.S. Iyengar, Louisiana State University

Researchers are using a sensor-input-based metric to develop a team of robots that would have the capability to learn their roles and improve strategies so that they can meet their overall goals in dynamic unstructured environments such as underwater or urban settings in which communications and monitoring are difficult.

For a robot to operate autonomously in a dynamic environment, it must be capable of adapting itself without the help of humans. The ultimate goal of our research is to provide teams of unmanned underwater vehicles (UUVs) some of the abilities of animals to adapt to their environment using their memories, without requiring exhaustive trial-and-error testing or complex modeling of the environment.

We focus on UUVs because they offer the promise of making dangerous tasks such as searching for underwater hazards or surveying the ocean bottom more safe and economical for government and commercial operations. We adopt a team concept to reduce overall mission cost using several low-cost subordinate UUVs to augment the sensor capabilities of a higher-capability lead UUV. Our goal is to develop a team of robots that would have the capability to learn their roles and improve team strategies so that the team can meet its overall goals in dynamic unstructured environments such as underwater or urban settings in which communications and monitoring are difficult.

Our research uses a sensor-input-based metric for success combined with a training regimen based on recently collected memories—a temporal series of sensor/action relationships—in which robots with "ears" listen for a leader robot and attempt to follow,[1] and where the ensuing formations are a result of emergent behavior.[2] For this application, the sensor input is the sound intensity in the left and right ear, and the action is to turn left, go straight, or turn right, keeping the intensity within a certain range in both ears.

UNDERWATER NAVIGATION

One of the greatest challenges for underwater autonomous operations is navigation. Currently, most precision underwater navigation relies on some sort of external infrastructure such as surface ships or underwater beacons placed in known positions. Subsea navigation uses these assets as reference points. However, this limits the operation of UUVs to fairly small areas, and some situations require assessing an area's environmental or commercial attributes before an infrastructure exists. To accomplish such tasks, the UUV team must be able to navigate to an area, carry out its task, and return, requiring expensive and complex navigation systems. Seawater's varying physical properties, along with acoustic issues such as spreading, reverberation, and multipath, make autonomous, nonifrastructure-based underwater navigation a difficult task.Any mission involving multiple UUV's relies on their capability to navigate as a team. A typical mission will have many distinct phases, requiring the smooth transition between formations. Initially, the UUVs will be onboard a host vessel. After the UUVs have been sea- and mission-prepped, they will be put into the water and will form into a group to travel to the area of interest. The current assumption is that at least one vehicle will have an advanced positioning system on board and that the others will navigate relative to this vehicle.

Published by the IEEE Computer Society

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2006**	2. REPORT TYPE	3. DATES COVERED **00-00-2006 to 00-00-2006**
4. TITLE AND SUBTITLE **Memory-Based In Situ Learning for Unmanned Vehicles**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Louisiana State University,Department of Computer Science,Baton Rouge,LA,70803**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**5**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Upon arriving at the area of interest, the UUVs will change into task-specific formations and execute their mission-related goals. When the mission is completed, the UUVs will journey back to their host vessels, where the collected data will be processed and disseminated.

Figure 1 shows a team of five UUVs moving in a V-formation. The red UUV is the leader, while the green UUVs are followers. In this illustration, the follower UUVs are assisting the leader by extending its sensor footprint, indicated by the yellow ovals; overall costs are reduced because the green UUVs do not need the same capabilities (long-range communications and navigation) as the red UUV.

The leader/follower concept uses relative navigation between vehicles to provide an attractive and effective solution—a solutoin that is used abundantly in nature and in human activities involving multiple vehicles. However, because of the undersea environment's constantly changing properties—currents, density, bottom composition, biofouling, and so forth—and different mission-specific payloads, the vehicles must be able to adjust their control strategies.

Our work recognizes these factors and is built upon earlier work that simulated formations of neural-network-controlled vehicles and then was extended to wheeled mobile robots using acoustic sensor systems.[1] We focus on memory-based learning algorithms designed to reduce the number of trial solutions (environmental exploration) the robotic system requires. The anticipated benefits include reduction of time for setup and calibration of sensor systems and, in the context of a feedback-based robot architecture, quicker adaptation to changing environments.

Our work is distinctive because it uses machine-learning techniques to learn the control laws to move the UUVs into (acquire) formation, and maintain (follow) it. The system uses machine-learning techniques to learn the low-level sensor/action relationships, while it uses emergent behavior to construct the formations.[2] Our research uses data acquisition methods to generate controllers without a priori knowledge or physical repetition of candidate solutions.

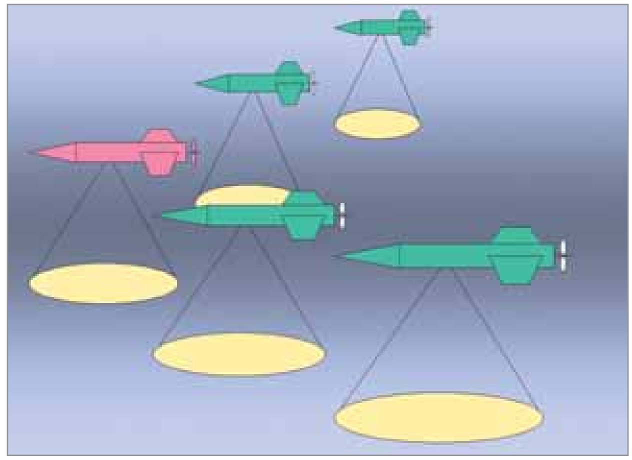


Figure 1. UUV team. In this team of five UUVs, the red UUV is the team leader. In this leader/follower team concept, the red leader UUV contains a high-accuracy navigation system, while the green follower UUVs navigate relative to the leader by acoustically sensing its position.

MEMORY-BASED LEARNING

If robots can learn from recent memory, researchers can avoid directed testing of trial solutions. By recording sensor data/action pairs and actions that optimize goals, a robot can create a "draft" controller that researchers can iteratively improve as the robot operates in its environment.

Our *random-but-purposeful controller* uses sensor feedback to continue actions that move it closer to its goal. As long as the feedback indicates that the robot is meeting its goal, it continues what it is doing; if not, it randomly chooses another action. Since the selection of the new action is random, examples of correct actions are distributed relatively evenly. In the case of a follower robot exploring its environment, the left, right, and straight examples will be distributed relatively evenly.

Figure 2 shows a graphic example of the robot trajectory using the random-but-purposeful training controller. In some tests, good examples consisted predominately of either left or right turns; as a result, the generated neural-network controller did not learn how to turn in both directions. To contend with this deficiency, we used a mirroring technique that generated "extrapolated" memories where a left turn closer to the source generated an identical right turn to complete the training set. These training sets were then passed to a simple genetic–algorithm training process that generates a feed-

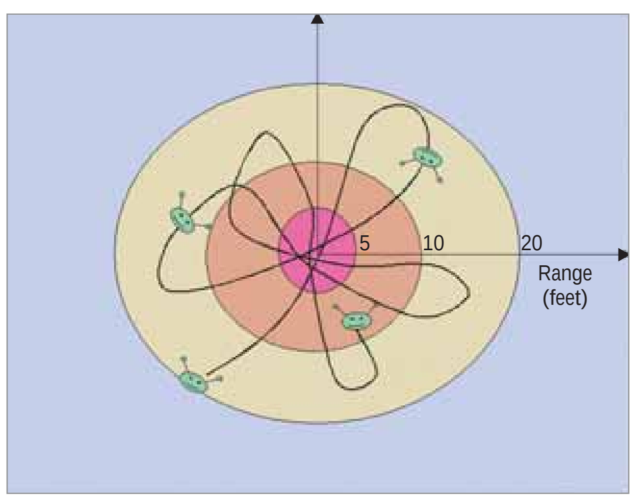


Figure 2. Random-but-purposeful controller. In this example path, the sound source is at the center of the plot and the smiley faces with antenna denote the robot and its orientation as it follows the path.

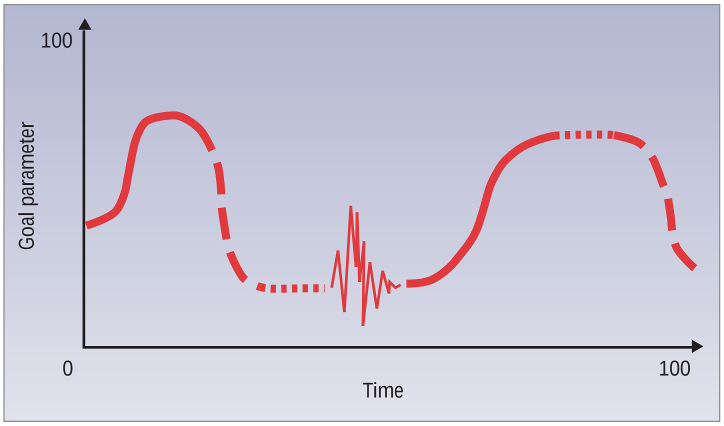


Figure 3. Training set. The solid part of the line fits in the category of rising parts, the dashed part in the dropping category, the dotted part in remaining stable category, and the thin part near the middle of the time series is the fluctuating category.

forward, neural-network controller. Learning from data collected directly from the operation environment lessens calibration time for sensors and equipment.

REACTIVE CONTROLLER

For a reactive controller, the actions at time *t* are based solely upon the current sensor readings. Consequently, the reactive-learning algorithm uses a training set comprised strictly of individual sensor/action pairs that are marked good if the action results in an increased sensor value or bad for a decreased value. The controller examines the recent memory generated by the random-but-purposeful exploration algorithm for occurrences of these examples.

In this process, the positive examples are used to form a set of training examples used by a genetic algorithm to train a neural network. An intensity filter ensures that actions have a direct effect on sensor results. Once this cause-and-effect relationship is established, the system sorts the examples into positive and negative sets. The work presented here uses only the positive examples; however, simulations have demonstrated that negative examples can also be used by rewarding a neural network if it takes a different action than that associated with the example.

One known drawback of this type of controller is that it cannot react to trends that occur over time. A classic example for the formation maneuvering task is the forward-backward ambiguity: If the amplitude in both ears remains the same, a following robot with a reactive controller cannot discern whether it is heading directly away from or toward the leader; without an additional observer to detect this trend, a robot could wander away from its leader in the wrong direction.

TREND BASED CONTROLLER

In the trend-based controller, the number of past inputs dictates the span of time over which the controller will be able to observe trends and also its ability to react quickly to sudden changes. The motivation for this type of controller was to ensure that it could properly handle the forward-backward ambiguity.

As Figure 3 shows, to create a training set for this controller, we split the recent memory into four categories of sensor-intensity trends: rising (R), dropping (D), stable (S), and fluctuating (F). The training sets are formed from the trend categories. For the rising and stable groups, the action taken during each particular run is assumed to be correct, so the neural networks in training are rewarded when they take this same action. However, when the goal parameter is dropping, the correct action is defined as the action in an adjacent or overlapping rising run because it is that action that alters the negative energy gain. If there is no adjacent or overlapping rising run, that particular dropping run is not used in the collection of training examples. The system currently does not use fluctuating runs to create training examples because there is no programmatic method of determining what the correct action is for this state.

EXPERIMENTAL RESULTS

Simulations have tested line formations with up to 30 robots. We also have used up to eight robots to test binary tree formations using attraction and repulsion. To physically simulate UUVs in the underwater environment, we used ActivMedia land robots equipped with audio transmission and "two-ear" listening systems operating semipassively, meaning that the robots do not exchange position or bearing and range information. Instead, each robot listens for a chirp emitted by its leader and steers itself toward the sound by turning in the direction of the strongest signal (left or right). The system uses a frequency multiplexed communication scheme in which each leader transmits in its own prespecified frequency band and followers are assigned to a leader by listening in the specified band. Followers use the signaling to determine a relative direction to steer toward the leader.

In quantitive testing using a computer simulation,[3] the reactive controller performed better but could not solve the forward/backward ambiguity, whereas the trend-based controller could under some conditions.

Figure 4 shows successful formation-maneuvering tests with a line of three robots using reactive feed-forward neural networks. Significantly, the individual robots are not aware of the concept formations; instead, they independently follow their assigned leaders based on the relative strengths of the chirps received at their microphones. The follower robots avoid colliding with their leaders by slowing down when sound intensity goes above a threshold, indicating that it is very close to the sound source.

Figure 4. Formation maneuvering. Time lapse photos show three robots using their "ears" to follow each other in the lab.

The formations are the result of emergent behavior, a global form of behavior that results solely from local, or bottom-up, activity. The advantage to using techniques based on these ideas is that they do not require a central controller, thus saving communications bandwidth, increasing robustness, and reducing internal system complexity.

Employing two learning algorithms helps to contend with the challenges presented by dynamic unstructured environments such as those found under water or in an urban environment where central control and planning is difficult for classic sense/plan/act systems. Learning what was done correctly in short exploration periods keeps the amount of required a priori knowledge to a minimum. The formations described in these tests result from emergent behavior in that no single robot is programmed with the concept of a line or other formation. Emergent behavior has been used to explain formations of birds and fish[2] and is a valuable tool for creating complex interactions among many individual entitieswithout relying on a centralized control scheme.

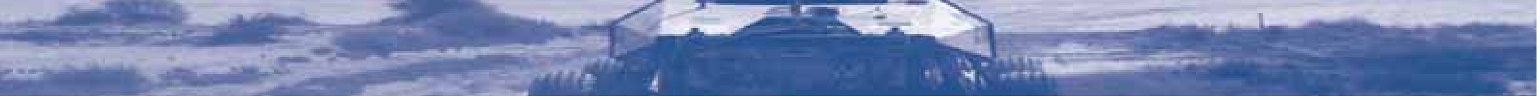
Acknowledgments

The authors thank NRL Positioning Navigation and Timing team members Marvin Roe, Donald Brandon, and Dustin Marmalich for their continual hard work during this project. This work was funded by the Office of Naval Research through the Naval Research Laboratory. The mention of commercial products or the use of company names does not in any way imply endorsement by the US Navy. Approved for public release; distribution is unlimited.

References

1. P. McDowell, B.S. Bourgeois, and S.S.Iyengar, "Formation Maneuvering Using Passive Acoustic Communications," *Proc. IEEE Int'l Conf. Robotics and Automation*, vol. 4, IEEE Press, 2004, pp. 3843-3848.
2. E. Thro, *Artificial Life Explorer's Kit*, Sam's Publishing, 1993, pp. 105-106
3. P. McDowell, *Biologically Inspired Learning System*, doctoral dissertation, Dept. of Computer Science, Louisiana State Univ., 2005.

Patrick McDowell is an assistant professor at Southeastern Louisiana State University. His research interests include machine learning as it applies to robotics, artificial life, and quantum computing. McDowell received a PhD in computer science from Louisiana State University. He is a member of the ACM. Contact him at pm@selu.edu.



Brian S. Bourgeois is the head of the Position, Navigation, and Timing team at the Naval Research Lab at Stennis Space Center, Mississippi. His research interests include autonomous surveying, signal processing, and GIS applications. Bourgeois received a PhD in electrical engineering from Tulane University. He is a member of the IEEE. Contact him at bsb@nrlssc.navy.mil.

Donald A. Sofge is a computer scientist and researcher for the Navy Center for Applied Research in Artificial Intelligence at the Naval Research Lab, Washington, D.C. His research interests include autonomous systems, machine learning, computational intelligence, and quantum AI. Sofge received an MS in electrical engineering from the University of Florida and is a PhD student in information technology at George Mason University. He is a member of the IEEE Robotics and Automation Society. Contact him at don.sofge@nrl.navy.mil.

S.S. Iyengar is the chairman and Roy Paul Daniels Chaired Professor of Computer Science at Louisiana State University and is also the Satish Dhawan Chaired Professor at the Indian Institute of Science. His research interests include path planning, high-performance parallel and distributed algorithms, and data structures for image processing. Iyengar received a PhD in engineering from Mississippi State University. He is an IEEE Fellow and a member of the ACM. Contact him at iyengar@bit.csc.lsu.edu.